Letter

DOI: 10.5582/irdr.2025.01079

AI-driven enhancements in rare disease diagnosis and support system optimization

Xin Wang, Da He*, Chunlin Jin

Shanghai Health Development Research Center, Shanghai, China.

SUMMARY: Rare diseases are characterized by an extremely low prevalence, high phenotypic heterogeneity, and complex pathogenesis. This combination of factors presents significant challenges, including prolonged diagnostic delays, lack of standardized care, and difficulties in pathological interpretation. The integration of artificial intelligence (AI) offers a transformative approach to overcoming these barriers. In recent years, researchers worldwide have been actively exploring the use of AI to diagnose and manage rare diseases. Key advances include few-shot learning algorithms designed to tackle data scarcity, clinically validated foundation models that enhance diagnostic consistency across institutions, and multimodal AI frameworks that integrate imaging, genomic, and phenotypic data to improve diagnostic accuracy. In addition, there is growing recognition that AI can enhance diagnostic efficiency and thereby optimize support systems for rare diseases. As challenges such as AI model interpretability and data equity are addressed, AI is expected to make significant strides in the diagnosis and treatment of rare diseases.

Keywords: rare disease, artificial intelligence, diagnosis, support system

Rare diseases are characterized by a low prevalence, high phenotypic heterogeneity, and complex pathogenesis. Patients with these diseases face prolonged diagnostic delays averaging 4–5 years (*1*), insufficient standardization of care, and challenges in pathological interpretation. The lack of precise diagnostic and therapeutic technologies is a critical barrier to improving patient outcomes. The further integration of artificial intelligence (AI) in healthcare has the capacity to overcome data scarcity ("the curse of dimensionality") (*2*), integrate multimodal clinical data, and enhance diagnostic and therapeutic precision. The accelerating advancement of AI is driving its growing use in the diagnosis and treatment of rare diseases. Our goal in this Letter is to summarize recent progress and to offer actionable insights for clinical practice.

Innovation in diagnostic algorithms for few-shot learning scenarios

The scarcity of annotated data for rare diseases is a primary bottleneck that restricts the development of diagnostic models. The recently introduced RetiZero model, a knowledge-rich vision-language foundation model pre-trained on over 400 rare and common fundus diseases, demonstrates remarkable adaptability with minimal data (*3*). When fine-tuned with only five images per category, it achieved exceptional area under the curve (AUC) scores (0.967, 0.859, 0.942) across three clinical datasets, significantly outperforming existing models. RetiZero integrates a masked autoencoder (MAE) with contrastive language-image pre-training (CLIP) to enable robust feature learning from minimal data, offering a powerful paradigm for scenarios with limited data.

Further demonstrating the versatility of few-shot learning in rare disease diagnosis, Alsentzer *et al.* introduced SHEPHERD, a geometric deep learning model that leverages a biomedical knowledge graph and simulated patient data for phenotype-driven genetic diagnosis (*4*). SHEPHERD was trained on over 40,000 synthetic patients with 2,134 rare diseases. When evaluated on a real-world, independent cohort from the Undiagnosed Diseases Network (UDN), it achieved a top 1 accuracy of 40% in causal gene discovery from EXPERT-CURATED gene lists and effectively retrieved "patients like me." Its ability to generalize across clinical sites and novel diseases underscores the potential of knowledge-guided, few-shot learning frameworks to overcome data scarcity in genomics and accelerate the diagnosis of rare genetic conditions.

Diagnostic and therapeutic applications of clinical-grade foundation models

The development of foundation models has overcome the limitations of "data dependency" in diagnosing rare

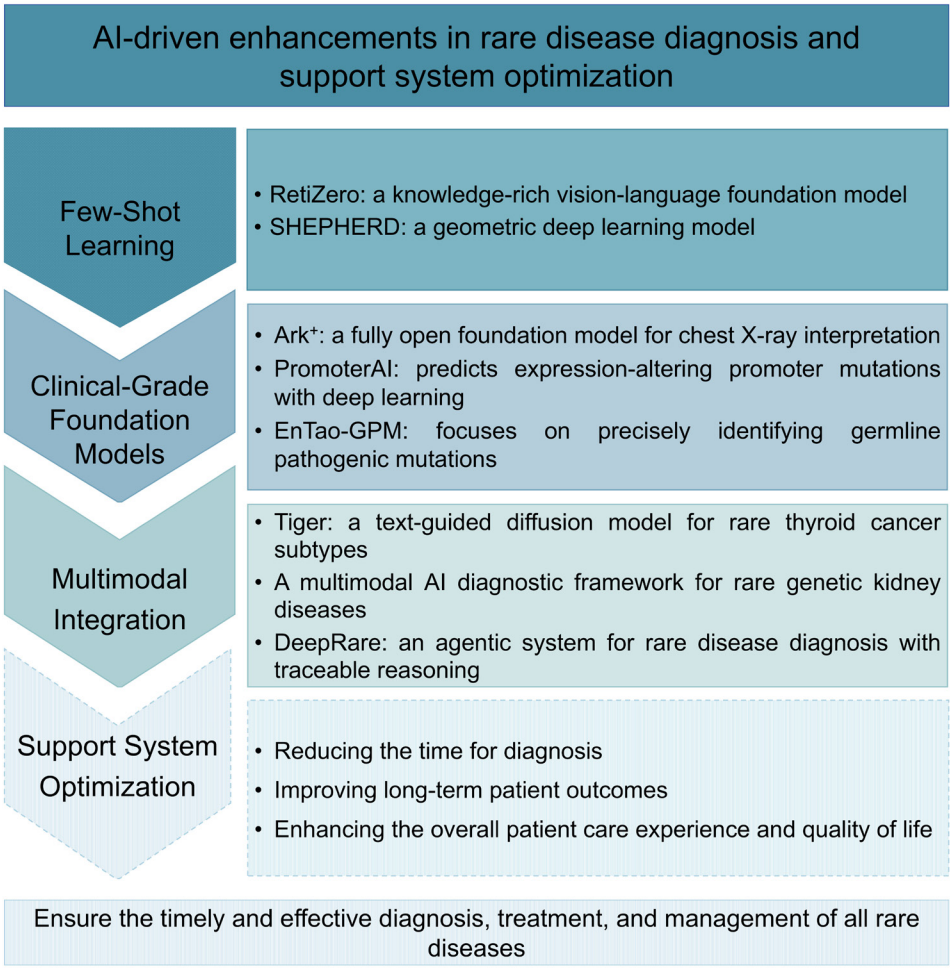

Figure 1. AI-driven enhancements in rare disease diagnosis and support system optimization. AI, artificial intelligence.

diseases. These models enable consistent interpretations across different clinical centers and disease types. Ark$^+$ is a fully open foundation model for chest X-ray interpretation, pre-trained by cyclically accruing and reusing knowledge from heterogeneous expert labels across six public datasets (*5*). It demonstrates strong generalizability and excels in few-shot learning to diagnose rare thoracic conditions, effectively addressing long-tailed disease distributions. Its capability for zero-shot transfer and federated learning enables robust, privacy-preserving diagnoses across institutions, making it a powerful, clinically adaptable tool for diagnosing both common and rare diseases.

Another example is PromoterAI, a deep neural network developed by Jaganathan *et al*. (*6*). First, it was trained on large-scale functional genomics datasets to learn the "regulatory syntax" underlying transcriptional and epigenetic programs. Then, it was fine-tuned using a curated set of rare promoter variants associated with outlier gene expression. By accurately detecting expression-altering variants, PromoterAI expands the clinical utility of genome sequencing and increases the yield of diagnoses for rare genetic diseases.

Similarly, Lin *et al*. developed the EnTao-GPM DNA Foundation Model, which focuses on precisely identifying germline pathogenic mutations (*7*). Through cross-species targeted pre-training, fine-tuning on the ClinVar database and the Human Gene Mutation Database, and integration with large language models to generate interpretable clinical reports, the model significantly improves the accuracy of mutation classification (achieving an AUC of 0.963 for single nucleotide variants). The model serves as a powerful tool for shortening the diagnostic process for rare diseases and guiding personalized treatment.

Precision diagnosis through multimodal data integration

The diagnosis and treatment of rare diseases frequently require integrating multi-source data, including medical imaging, genomic data, and clinical phenotypes. The ability of AI to seamlessly integrate diverse data sources has been found to enhance diagnostic accuracy. Dai *et al*. developed a text-guided diffusion model (the Tiger Model) for rare thyroid cancer subtypes (*8*). This model integrates ultrasound imaging features with clinical text descriptions. The generated synthetic images were validated by physician Turing tests, achieving a realism rate of 92.2%. This approach improved the AUC for

subtype classification from 0.7364 to 0.8442, thereby laying the foundation for the development of customized treatment strategies.

Chen *et al*. proposed a multimodal AI diagnostic framework for rare genetic kidney diseases (*9*). This framework integrates genomic data, human phenotype ontology (HPO) terms, and imaging features. The researchers discussed the potential of AI approaches to address significant underdiagnosis and shorten the diagnostic odyssey for these conditions, thereby improving diagnostic precision.

Zhao *et al*. developed the DeepRare system (*10*), which features a three-tier architecture of "central host–specialized agent servers–heterogeneous web-scale medical sources," supporting multi-source data input and traceable reasoning. The system's performance was evaluated on 6,401 cases from multiple global centers, achieving an average Recall@1 score of 57.18% with HPO term input alone. This was further enhanced to 70.6% in multi-modal scenarios. The system has been implemented in numerous domestic and international institutions.

Optimization of rare disease support systems

There is growing recognition that AI can enhance diagnostic efficiency and thereby optimize support systems for rare diseases. This perspective has been discussed at numerous academic conferences, including the Second "Hai Shang" Rare Disease Forum, held in Shanghai, China, on November 14-15, 2025, which focused on the theme of "AI-Powered Diagnosis, Protection, and Humanistic Care in Rare Diseases."

One can reasonably expect improved diagnostic efficiency — driven by AI — to systematically propel the optimization of rare disease support systems and the implementation of humanistic care, thereby establishing a virtuous cycle where "diagnosis and treatment, support, and humanistic care" mutually reinforce one another (Figure 1). First, AI-assisted diagnosis reduces the time patients spend on their "diagnostic odyssey," which is expected to reduce unnecessary testing and ineffective treatments. This approach thus has the potential to alleviate financial burdens on both patients and the healthcare system, while also helping to preserve scarce, quality medical resources. In addition, the precise identification of pathogenic variants by AI provides a critical pre-symptomatic intervention window for hereditary rare diseases, thereby improving long-term patient outcomes. Finally, AI tools support primary care physicians by providing expert-level diagnostic guidance, minimizing misdiagnoses caused by limited experience. This, in turn, reduces patient suffering and enables more personalized treatment regimens, ultimately enhancing the overall patient care experience and quality of life.

In conclusion, as challenges related to the interpretability of AI models and data equity are addressed, the field of AI is poised to make significant advances in the diagnosis and treatment of rare diseases. The goal is to move from a focus on "precision" to a more comprehensive approach, aiming for "universal benefit." This will involve the development of tools and methodologies that ensure timely and effective diagnosis, treatment, and management of all rare diseases.

Funding: None.

Conflict of Interest: The authors have no conflicts of interest to disclose.

References

1. Marwaha S, Knowles JW, Ashley EA. A guide for the diagnosis of rare and undiagnosed disease: Beyond the exome. Genome Med. 2022; 14:23.
2. Banerjee J, Taroni JN, Allaway RJ, Prasad DV, Guinney J, Greene C. Machine learning in rare disease. Nat Methods. 2023; 20:803-814.
3. Wang M, Lin T, Lin A, *et al*. Enhancing diagnostic accuracy in rare and common fundus diseases with a knowledge-rich vision-language model. Nat Commun. 2025; 16:5528.
4. Alsentzer E, Li MM, Kobren SN, Noori A, Kohane IS, Zitnik M. Few shot learning for phenotype-driven diagnosis of patients with rare genetic diseases. NPJ Digit Med. 2025; 8:380.
5. Ma D, Pang J, Gotway MB, Liang J. A fully open AI foundation model applied to chest radiography. Nature. 2025; 643:488-498.
6. Jaganathan K, Ersaro N, Novakovsky G, *et al*. Predicting expression-altering promoter mutations with deep learning. Science. 2025; 389:eads7373.
7. Lin Z, Sun H, Guo Y, Yang Y, Wang Y, Hu B, Ye C, Yang Q, Zhong F, Zhang X, Liu L. EnTao-GPM: DNA foundation model for predicting the germline pathogenic mutations. arXiv [Preprint]. 2025; arXiv:2507.21706.
8. Dai F, Yao S, Wang M, *et al*. Improving AI models for rare thyroid cancer subtype by text guided diffusion models. Nat Commun. 2025; 16:4449.
9. Chen X, Burgun A, Boyer O, Knebelmann B, Garcelon N. Artificial intelligence and perspective for rare genetic kidney diseases. Kidney Int. 2025; 108:388-393.
10. Zhao W, Wu C, Fan Y, Zhang X, Qiu P, Sun Y, Zhou X, Wang Y, Sun X, Zhang Y, Yu Y, Sun K, Xie W. An agentic system for rare disease diagnosis with traceable reasoning. arXiv [Preprint]. 2025; arXiv:2506.20430.

Received November 18, 2025; Revised November 25, 2025; Accepted November 27, 2025.

**Address correspondence to:*
Da He, Shanghai Health Development Research Center, No. 602, West Jianguo Road, Xuhui District, Shanghai 200031, China.
E-mail: heda@shdrc.org

Released online in J-STAGE as advance publication November 28, 2025.